MATH11028 Simulation: Project 1

Lorenzo Gianelle, Xi Chen

1 Introduction

Our team has been assigned to evaluate the performance of BoxCar, a ride-sharing company operating in the city of Squareshire. BoxCar operates through a system that matches drivers with riders.
To assess the efficiency of BoxCar, a simulation model has been developed, aimed at evaluating key performance indicators related to both customer and driver satisfaction.
The project begins by modelling the system using probabilistic distributions provided by BoxCar, in order to get initial information into its performance. Next, using data provided from the company, the distributions have been tested, to evaluate if they accurately reflect the real-world situation. If discrepancies arise, alternative distributions are proposed. Finally, the simulation is run again, results are analysed, and potential improvements are identified and implemented to optimize BoxCar's system.

2 Method

To evaluate the performance of BoxCar's system, a simulation model has been developed to track the functioning of BoxCar in Squareshire. This model allows the analysis of key performance indicators related to the satisfaction of both drivers and passengers.

2.1 Hypothesis

BoxCar has provided assumptions, including probability distributions and a map of Squareshire, to help perform the desired simulation.

2.1.1 Squareshire

Squareshire is represented as a 20-mile by 20-mile square. Coordinates will be used to track locations of drivers and riders within the city, during the simulation. For consistency, the (0,0) coordinate has been assigned to the bottom left corner of the city.

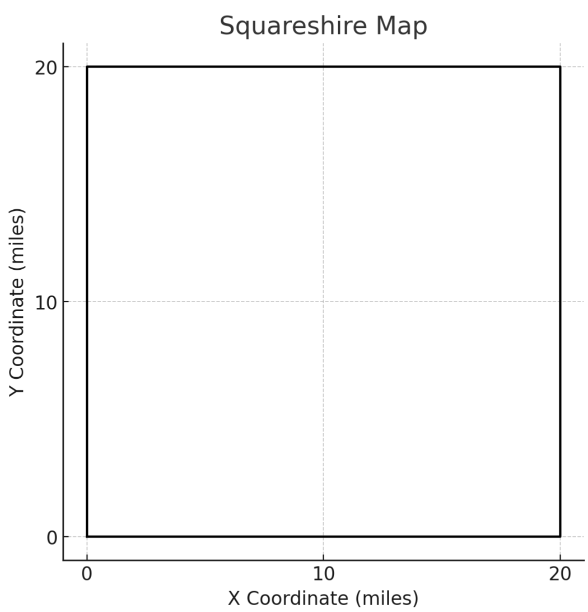

Figure 1: Map of Squareshire

2.1.2 Drivers

- Drivers become available at random inter-arrival times, following an exponential distribution with a rate parameter $\lambda = 3$ hours (average arrival rate of 3 drivers per hour).
- Each driver's working time is randomly determined at the start of the shift and follows a uniform distribution between 5 and 8 hours. If a driver is still on a ride when their shift is supposed to end, they complete the ride before going offline.
- Drivers can start their shift from any point of the city, meaning their initial location is distributed as a Uniform over a two-dimensional region ($[0, 20]$ x $[0, 20]$). Until a rider is assigned, the driver remains at the same location.

2.1.3 Riders

- Riders appear at random times following an exponential distribution with a rate of $\lambda = 30$ riders per hour.
- The pick-up and drop-off locations are independent of each other and are uniformly distributed across Squareshire.
- Each rider has a limited patience time, which follows an exponential distribution with a rate $\lambda = 5$ per hour. If a rider is not matched with a driver within this time, they cancel the request and leave the system. However, the ride does not need to start within this period, as long as the rider is informed that a driver is on the way, they remain in the system until the trip is completed.

2.1.4 Rider - Driver Match

When a rider requests a ride through the app, the system checks for available drivers who are not currently serving a passenger. If there are available drivers, the app assigns the closest driver to the rider. The driver then begins travelling toward the rider's pick-up location.
Once the driver reaches the rider, they start the trip toward the drop-off destination. The total trip consists of two parts:

- The pick-up ride: the driver travels to reach the rider.
- The drop-off ride: the driver travels from the rider's pick-up location to the destination.

The travel time, for each part of the ride, depends on the Euclidean distance (d_{OD}) between the two locations. The average speed is assumed to be 20 miles per hour, so the expected trip duration is calculated as $\mu_t = 20/d_{OD}$. However, to reflect real world variations (for example, traffic along the way), the actual trip time follows a uniform distribution between $0.8\mu_t$ and $1.2\mu_t$.
At the end of the ride, the rider pays the driver a base fare of £3, plus £2 per mile travelled from the pick-up location to the final destination. The driver, however, must cover fuel costs of £0.20 per mile, which applies to all miles driven, including those spent travelling to pick up the rider.
At this point, after completing the ride, if drivers have not ended their shift and there are other passengers waiting, they are immediately assigned to the closest available rider and begin the process again. If no passengers are waiting, the driver remains available at the same location until a new ride request appears.

2.2 Simulation Structure

The simulation implemented operates as a discrete event simulation, meaning that the system evolves over time t based on events that occur sequentially. The implementation has been developed using a procedural approach in R.
At the core of the simulation is the event calendar, which stores all scheduled events in chronological order. The simulation processes events one at a time, updating the system state and scheduling new events. At the end of each iteration, the event calendar is sorted to ensure that the next event that occurs is at hand.
To initialize the simulation, empty datasets are created for drivers, passengers, and events. The first driver and passenger arrivals are scheduled based on their random inter-arrival times, creating the initial state of the system. From this point, the simulation starts, it does not progress in fixed time steps, but

instead jumps forward to the exact moment when the next event occurs.
As the simulation runs, it records key performance indicators (KPIs) to assess system efficiency.

2.2.1 Agents of the simulation

The system consists of two key agents:

- Drivers: become available, wait for matching, pick up passengers, complete trips, and go offline after their working shift.
- Passengers: request rides, wait for matching, wait for pick-up, complete trips, or leave the system if they wait too long without being assigned a driver.

2.2.2 System States

At any given time, the system can be in different states, depending on the availability of drivers, request of riders and state of ongoing matches. Drivers can be in one of the following states:

- Available: the driver is waiting to be assigned to a passenger.
- Busy to pick-up :the driver has been assigned a passenger and is travelling to the pick-up location.
- Busy to drop-off: the driver has picked up the passenger and is driving toward the drop-off location.
- Offline: the driver has completed the shift and is no longer available for new rides.

On the other hand, riders can be:

- Waiting: the rider has requested a ride and is now waiting for a match with a driver.
- Waiting for pick-up: a driver has been matched to the rider, who is now waiting for the car to arrive.
- Travelling: the rider is travelling toward the destination.
- Dropped Off: the passenger has reached the destination and exited the system.
- Abandoned : the rider left the system after waiting too long without being matched to a driver.

During the simulation, events such as driver arrivals, passenger requests, pick-ups, and drop-offs modify the system state.

2.2.3 Events that change the system state

The system evolves over time through a series of discrete events, each of which changes the status of drivers and riders.
The key events that modify the system state are:

- Driver Arrival (DA): a new driver enters the system and becomes available at a location, if there are passengers waiting for a ride, the driver is immediately assigned to the closest one. A Driver Departure (DD) event is scheduled, defining the driver's shift duration.
- Driver Departure (DD): a driver, who is already in the system, goes offline after completing the shift. If the driver is currently transporting a passenger, finishes the ride before leaving the system. If the driver is available, simply exits the system.
- Passenger Arrival (PA): a new passenger enters the system and requests a ride from the pick-up location. If a driver is available, the nearest one is matched immediately and a Passenger Pick-Up (PP) is scheduled. If no driver is free, the passenger waits, and a Patience Limit (PL) event is scheduled to track how long the rider is willing to wait before abandoning the system.
- Passenger Leaves (PL) : happens if a passenger is not assigned a driver before the patience time is over: the rider exits the system.
- Passenger Pick-Up (PP) : the driver reaches the pick-up location and starts the ride. A Passenger Drop-Off (PD) event is scheduled based on the trip duration.

- Passenger Drop-Off (PD): the couple driver-rider reach the drop-off location. If the driver has ended the shift, leaves the system, otherwise if there are waiting passengers, the driver is immediately assigned to the closest one and a new PP is scheduled. If no passengers are waiting, the driver remains available until a new request appears.
- Termination (T): the simulation ends after five days and performance metrics can be analysed.

2.2.4 Key information of drivers and riders

In the simulation, key information about drivers and riders are stored to implement the coding.

- Drivers' Dataset
  - **index.d**: the index of the driver.
  - **location.d**: the current location of the driver in the city. This will be updated when the driver picks up or drops off a rider.
  - **DA**: the time when the driver goes online.
  - **DD**: the time when the driver goes offline.
  - **DB**: indicates whether the driver is busy or not. If DB = FALSE, the driver is available for a ride. If DB = TRUE, the driver is either busy or already offline and cannot be selected.
  - **income**: the total income of the driver. When the driver goes to pick up a rider, the income decreases because of the cost of fuel. When the driver drops off a rider, the income increases.
  - **working.time**: the total working time of the driver, including the time spent picking up the rider and driving to the destination.
  - **single.drive.t**: stores the time of the current drive performed, such as picking up a rider or driving to the destination.
  - **money.drive**: Similar to **single.drive.t**, but it stores the expected cost or income of the ride.
- Riders' Dataset
  - **index.p**: the index of the rider.
  - **location.begin**: the location where the rider appears.
  - **location.end**: the location of the rider's destination.
  - **AB**: indicates whether the rider has abandoned. If AB = TRUE, the rider has abandoned the system.
  - **PA**: the time when the rider appears.
  - **PB**: indicates whether the rider is matched to a driver. If PB = TRUE, the rider is matched and cannot be selected again.
  - **waiting.time**: the time duration from when the rider appears to when they are picked up.

2.2.5 Key Performance Indicators

The KPIs metrics measured help assess system performance: rider and driver satisfaction.
For drivers, the simulation measures:

- Income per hour: average earnings that drivers accumulates over their shift, made by base fares plus per-mile charges.
- Resting time: total time that a driver spends waiting for a new passenger during a shift.

In order to measure fairness among drivers, standard deviations and quantiles of the KPIs will be computed.
On the other hand, for riders the simulation measures:

- Waiting time: the time each passenger spends waiting for a driver after requesting a ride.
- Abandonment rate : the percentage of passengers who cancel their request due to waiting times exceeding the patience time.

The KPIs provide insights into the balance between driver supply and rider demand, helping to identify inefficiencies such as long waiting times, too low or too high driver employment or high abandonment rates. These indicators will result useful, since through their analysis it will be possible to suggest and then implement changes in the structure of the model, with the goal of improving the performance of BoxCar's system.

2.2.6 Matching Driver and Rider

The matching process is critical in determining how quickly passengers are picked up and how efficiently drivers are utilized. There are two key matching functions used in the performed simulation:

- Matching Driver to Passenger (match.D.to.P): when a driver becomes available, they look for the nearest waiting passenger. The function calculates the Euclidean distance between the driver's current location and all available passengers. The passenger with the shortest distance is then selected.
- Match Passenger to Driver (match.P.to.D): when a passenger requests a ride, the system looks for the nearest available driver. The function calculates the distance between the passenger and all free drivers. The closest driver is assigned to pick up the passenger.

2.2.7 Drivers' income and cost

In the simulation, as stated before, each ride consists of two phases for the driver. The first step is when the driver travels from their current location to the passenger's pick-up point. The second phase is the actual ride, where the driver transports the passenger from the pick-up to the drop-off location. While both segments contribute to the cost of fuel for the driver, only the second part generates revenue, as passengers are not charged for the distance the driver travels before picking them up.
To track the impact of each trip on the income for the drivers, the simulation includes two separate functions:

- Cost of travelling to the pick-up location (driving.time.cost): before picking up the passenger, the driver incurs fuel costs for travelling to the pick-up location. This cost is calculated based on the Euclidean distance between the driver's starting position and the passenger's pick-up location, multiplied by the fuel cost of £0.20 per mile.
- Fare and cost calculation for the actual trip (driving.time.price): once the driver picks up the passenger, the system calculates the total fare based on the distance from the pick-up point to the drop-off location. The passenger is charged a base fare of £3, plus £2 per mile travelled. The driver still incurs a fuel cost of £0.20 per mile, which is deducted from their total earnings.

2.3 Improving the simulation

After implementation of the original simulation, a key issue has been observed: certain drivers had been unable to secure a match with riders, in particular they had been positioned at the edges of the city and had spent there their whole shift. In the previous approach, the matching process was based only on proximity between driver and rider, which could lead to unfair earnings among the drivers. Drivers located in external areas could remain available for long periods, leading to earning little or no income. In order to increase fairness, the driver-passenger matching process has been updated, prioritizing drivers who have been working less. These adjustments aim to optimize driver satisfaction.

- Matching Passengers to Drivers (match.P.to.D): when a rider requests a trip, the system evaluates available drivers using a scoring function that considers both distance and work history. The score for each driver is calculated as:

  $$Score = Distance + Weight \times WorkingTime$$

  The driver with the lowest score is assigned to the rider. This ensures that drivers who have been available for longer periods are given a higher chance of securing a trip. Increasing the weight, means giving more priority to drivers who have worked less.

In Section 4, the changes will be evaluated and discussed in more depth.

3 Testing

To ensure the accuracy of the simulation, the distributions provided by the BoxCar's team have been analysed, using real-world data, always provided by the company. The objective has been to assess whether the assumed probability distributions used in the simulation accurately represent the reality. This step is important because it ensures that the simulation is based on real-world conditions rather than assumptions that may not reflect what is actually happening. If the distributions used in the model do not match real data, the simulation results could be misleading, making any conclusions unreliable.

3.1 Method for testing

To evaluate the goodness of fit, Kolmogorov-Smirnov (KS) tests and Chi-Squared tests have been conducted.

- The KS test compares the empirical cumulative distribution function of the observed data with the assumed theoretical distribution.
- The Chi-Squared test works through binning continuous distributions, measuring discrepancies between the observed and expected frequencies. The number of bins in the Chi-Squared test was determined using a standard rule of thumb, ensuring that each bin contains at least five expected observations, preventing bins with low counts. Also, in the testing, the bins were constructed to ensure equal probability for an observation to fall into each bin. This means that, for each distribution, the expected frequency of observations in each bin is constant. This method allowed for good flexibility, as it made it easier to adjust the test for different distributions.

All statistical tests will be conducted using a 90% confidence level, meaning that the significance threshold is $\alpha = 0.1$.
For simplicity, it has been assumed that the X and Y coordinates of the driver and rider locations are independent. This assumption simplifies statistical testing and allows for separate verification of the distributional properties of each coordinate.
Additionally, plots, like histograms and density plots, have helped, allowing to visualise the distribution of the data.
If significant differences have been identified between the observed and assumed distributions, alternative distributions have been proposed to improve the model's accuracy.

3.2 Testing Results

The analysis has been conducted using two datasets, one for drivers and one for riders, both imported into R for processing. The driver dataset contains information on driver starting location, arrival and offline time. The rider dataset contains information on request time, pick-up and drop-off locations and times, abandonment status. The data are collected on a time frame of 5 days (as the termination time set in the simulation). The number of driver in the system is 383, the number of riders is 3040.
Each dataset has been cleaned and structured to allow the analysis. In particular, location coordinates, which were originally stored in a single column, have been separated into two distinct columns: one for the X-coordinate and one for the Y-coordinate.

3.2.1 Driver Online Time

The online time of a driver indicates how long each driver has been in the system, it is computed from the information in the driver's dataset as the difference between the offline time and the arrival time. This difference is measured in hours to ensure consistency with the provided distribution.
The online time data have been compared with the assumed uniform distribution between 5 and 8 hours.

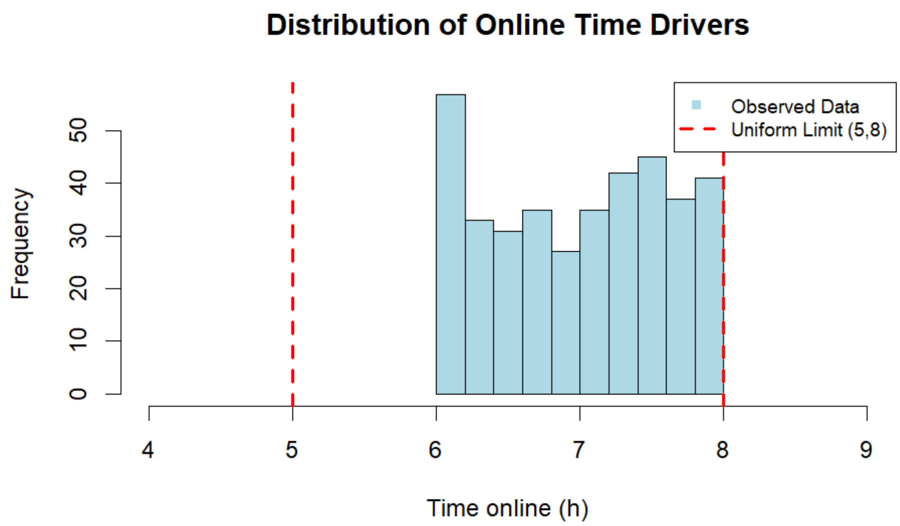

Figure 2: Distribution of Online Time for drivers and assumed uniform limits

The histogram in Figure 2 clearly shows that the assumed uniform distribution between 5 and 8 hours does not match the observed data. There are no recorded observations for driver online times between 5 and 6 hours, contradicting the initial assumption, instead, all drivers have online times between 6 and 8 hours.
Given this difference, the analysis is adjusted to focus on testing whether the observed data follows a uniform distribution between 6 and 8 hours.
It is assumed that each driver's shift duration is independent of others. This allows the use of a Chi-Squared test with the following hypothesis:

$$H_0 : \text{Online Driver Time} \sim \text{Uniform}(6, 8) \tag{1}$$

For the test, the number of bins is set to $k = 20$, resulting in an expected frequency of 19.15 observations per bin. The computed test statistic is $\chi^2 = 23.32$, while the critical value at a 90% confidence level is: $\chi^2_{(1-\alpha,k-1)} = 27.02$.
Since the test statistic is lower than the critical value, the test fails to reject H_0, meaning that the assumption of a uniform distribution between 6 and 8 hours is statistically valid.
This adjustment has an important impact on the simulation model, as it increases the average time drivers spend working. As a result, driver availability in the system will be higher than initially expected, which may influence other factors of the system performance.

3.2.2 Inter-Arrival Time for Drivers

The inter-arrival times of drivers have been computed as difference of consecutive arrival times of the drivers, then they have been analysed to assess whether they follow the assumed exponential distribution with rate $\lambda = 3$. A Chi-Squared test was performed under the assumption that driver arrivals are independent.
Initially, the test was conducted to verify whether the inter-arrival times follow an Exponential with rate $\lambda = 3$ distribution, using $k = 20$ bins with constant probability intervals. The test produced a test statistic of $\chi^2 = 44.07$ which is greater than the quantile value: $\chi^2_{(1-\alpha,k-1)} = 27.20$ at a 90% confidence level.
Since the test statistic is larger than the critical value, the test rejects H_0, meaning that the data does not support the assumption of an exponential distribution with rate 3.
Given this rejection, a new rate parameter λ was estimated using maximum likelihood estimation (MLE), giving:

$$\lambda_{\text{MLE}} = 4.09$$

A second Chi-Squared test has been performed to verify whether the inter-arrival times follow an Exponential with rate λ_{MLE}. The test statistic obtained was: $\chi^2 = 14.72$ which is lower than the critical value: $\chi^2_{(1-\alpha,k-2)} = 25.99$
The degrees of freedom used to compute the quantile are $k-2$, since a parameter has been estimated using the data. The test statistic is below the threshold, meaning that the test fails to reject H_0, confirming that the inter-arrival times of drivers follow an exponential distribution with rate $\lambda = 4.09$.

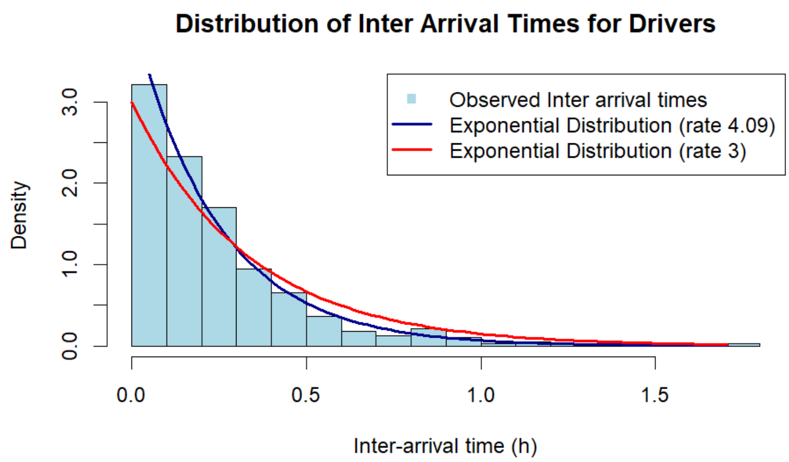

Figure 3: Histogram of observed inter-arrival times for drivers, with fitted exponential distributions for $\lambda = 3$ (red) and $\lambda = 4.09$ (blue).

This adjustment significantly impacts the simulation model, as it implies that drivers arrive more frequently than initially assumed: one more every hour than previously expected.

3.2.3 Initial drivers locations

The initial location distribution of drivers was analysed to determine if their starting locations follow a $\text{Uniform}(0, 20) \times \text{Uniform}(0, 20)$ distribution.

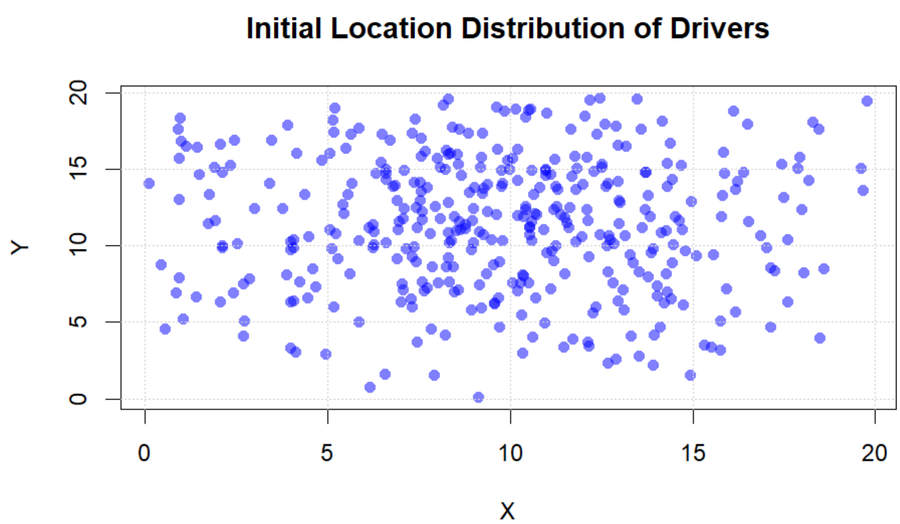

Figure 4: Initial Location of Driver in Squareshire

The scatter plot represents the observed locations, that do not look uniformly distributed on the Squareshire area. To test this hypothesis, a Chi-Squared test was performed by dividing the area into $k \times k$ grid cells (with $k = 5$), resulting in 25 bins.
The test statistic was computed as: $\chi^2 = 718.07$ which is much higher than the quantile value: $\chi^2_{(1-\alpha,k^2-1)} = 33.20$ meaning that the observed driver locations do not follow a uniform distribution across the city.
In order to find an appropriate distribution of the drivers' arrival location, the X and Y coordinates have been analysed separately, testing if they follow a truncated normal distribution over the range $[0, 20]$, using a Kolmogorov-Smirnov (KS) test. The parameters of the distribution for both the X and Y coordinate have been estimated using the MLE estimators.
For the X-axis:

$$\hat{\mu}_X = 9.61, \quad \hat{\sigma}_X = 4.78$$

The test statistic are:

$$D = 0.0355, \quad p\text{-value} = 0.7213$$

Since the p-value is greater than 0.1, the test fails to reject H_0, meaning that the X-coordinates can be modelled using a truncated normal distribution.
Similarly, the parameters for the Y-coordinate distribution are:

$$\hat{\mu}_Y = 11.83, \quad \hat{\sigma}_Y = 5.00$$

The test statistics:

$$D = 0.0277, \quad p\text{-value} = 0.9313$$

Again, the p-values is greater than 0.1, meaning that the test fails to reject H_0, confirming that the Y-coordinates follow a truncated normal distribution.
The drivers' arrival coordinates are no longer assumed to be uniformly distributed across Squareshire. Instead, certain areas will have a higher probability of driver arrivals, reflecting more realistic events.

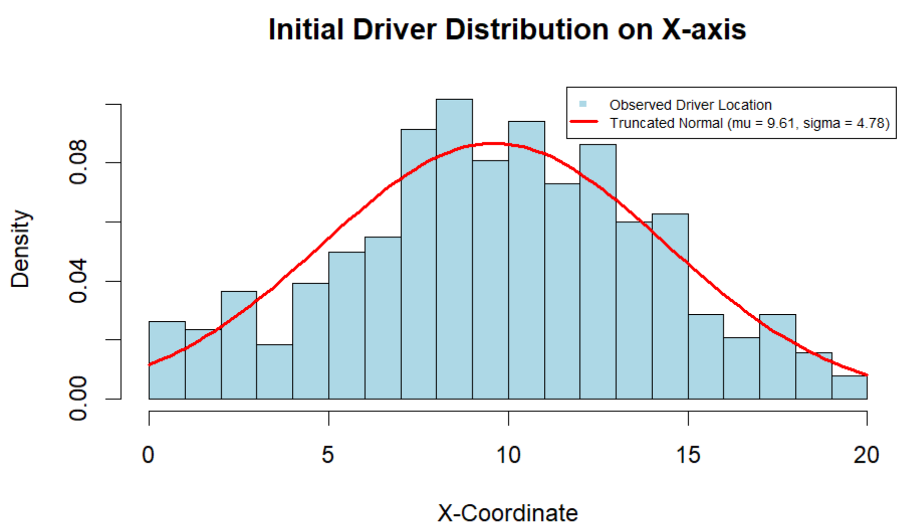

Figure 5: Observed data and Truncated Normal distribution for driver arrivals on the X-coordinate

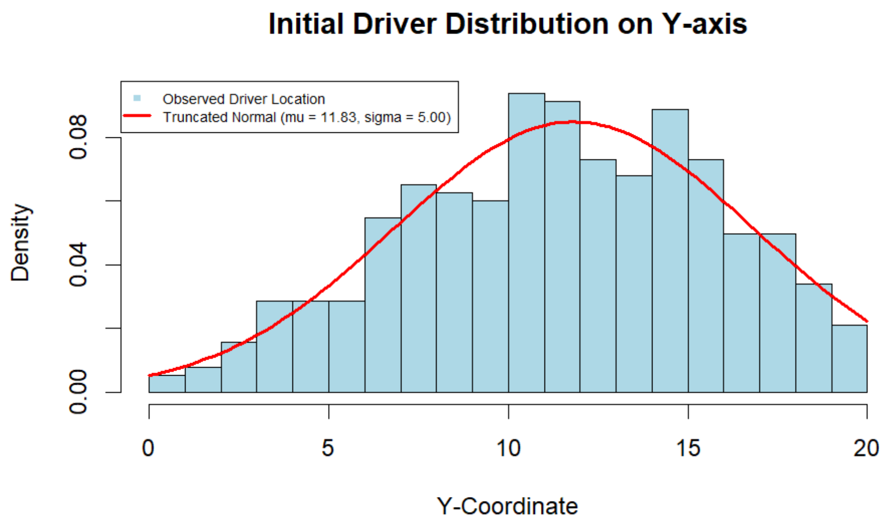

Figure 6: Observed data and Truncated Normal distribution for driver arrivals on the Y-coordinate

3.2.4 Rider Inter-Arrival Times

The inter-arrival times of riders have been analysed to verify whether they follow an exponential distribution with rate $\lambda = 30$. To test this assumption, a Chi-Squared test has been performed under the reasonable assumption that the timing of riders' requests are independent of each other. The test has been conducted using $k = 300$ bins, allowing the expected number of observations in each bin to exceed 10. The test statistic is: $\chi^2 = 288.15$ which is lower than the critical value at a 90% confidence level: $\chi^2_{(1-\alpha,k-1)} = 330.74$
The test fails to reject H_0, confirming that the arrival times of the riders follow an Exponential distribution of rate $\lambda = 30$.
This result validates the initial assumption made by BoxCar.

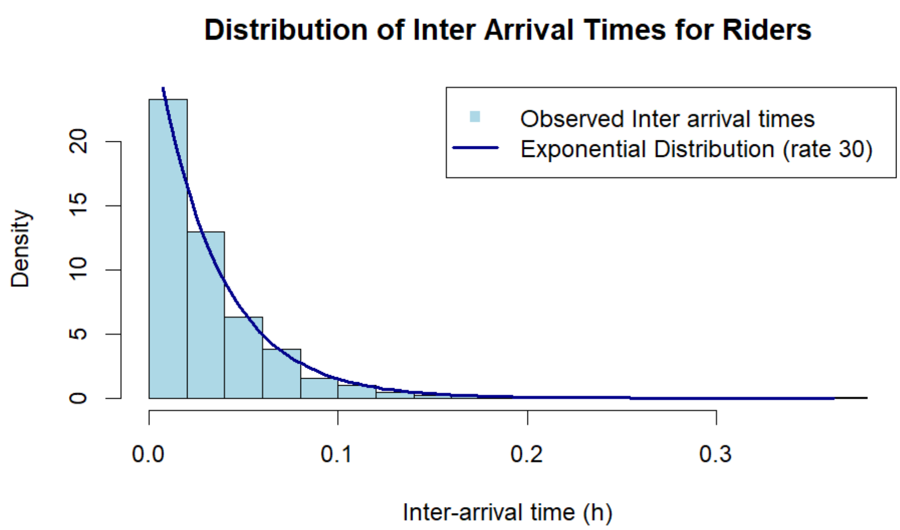

Figure 7: Histogram of observed rider inter-arrival times with tested Exponential(30) distribution.

3.2.5 Rider Pick-Up Location

The distribution of rider pick-up locations has been analysed to assess whether it follows a uniform distribution over the Squareshire area. The first test conducted was a Chi-Squared test, which strongly rejected the hypothesis that pick-up locations are uniformly distributed. This is evident simply by observing the spatial distribution across the city, represented in the next plot.

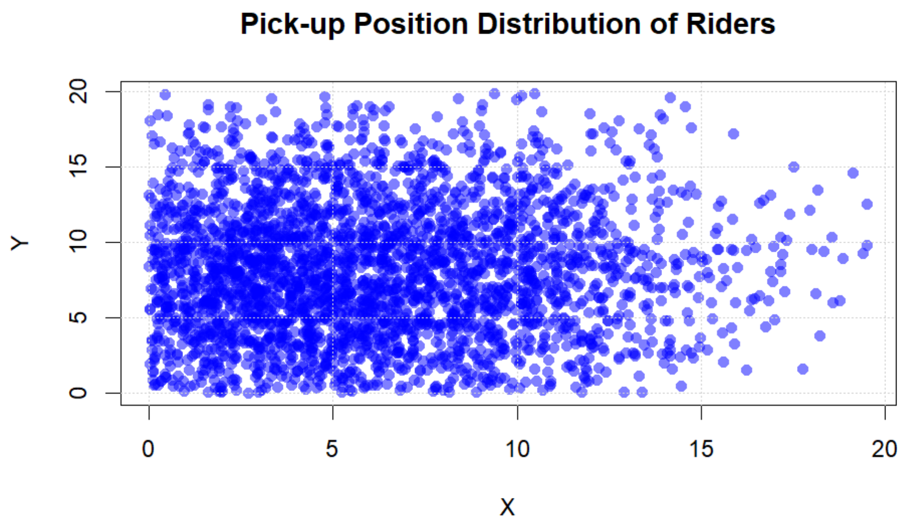

Figure 8: Scatter plot of observed pick-up locations.

The scatter plot clearly illustrates that the concentration of pick-up locations is skewed towards the left side of the city, indicating that ride requests are more frequent in that area. Riders tend to request rides more frequently in some areas than others, indicating that a uniform model is not appropriate.
Given this result, the next step is to model the X and Y coordinates separately using truncated normal distributions over the range $[0, 20]$, performing KS tests.
For the X-coordinate:
Using the MLE estimators for the parameters of the truncated normal distribution, yields:

$$\hat{\mu}_X = 4.52, \quad \hat{\sigma}_X = 5.16$$

The KS test has been performed to assess the distribution fit to the data, giving back results:

$$D = 0.0115, \quad p\text{-value} = 0.82$$

Since the p-value is crearly higher than 0.1, the test fails to reject H_0, meaning that the X-coordinates of pick-up locations are well approximated by a truncated normal distribution. Similarly, the truncated normal parameters for the Y-coordinate are:

$$\hat{\mu}_Y = 7.54, \quad \hat{\sigma}_Y = 5.09$$

The KS test for the Y-coordinate distribution resulted in:

$$D = 0.0094, \quad p\text{-value} = 0.9493$$

Again, the high p-value indicates that the Y-coordinates follow a truncated normal distribution.

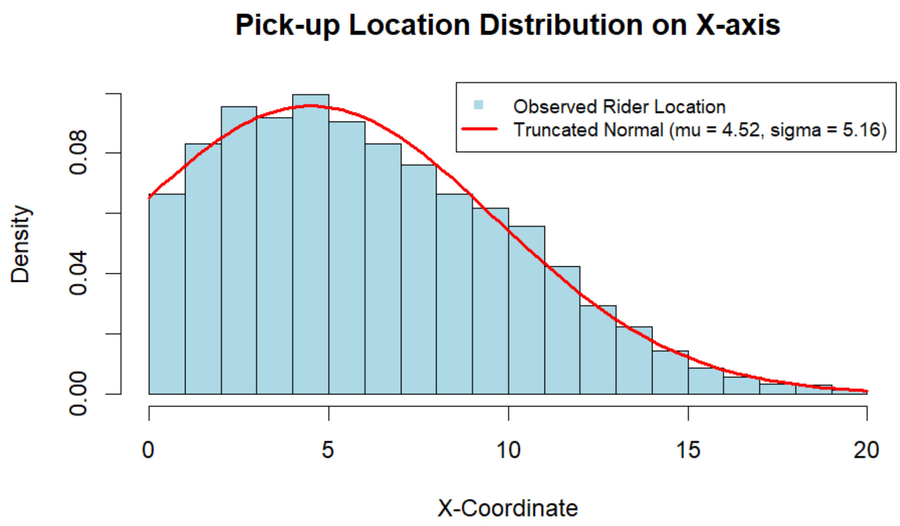

Figure 9: Observed data and Truncated Normal distribution for rider pick-up on the X-coordinate

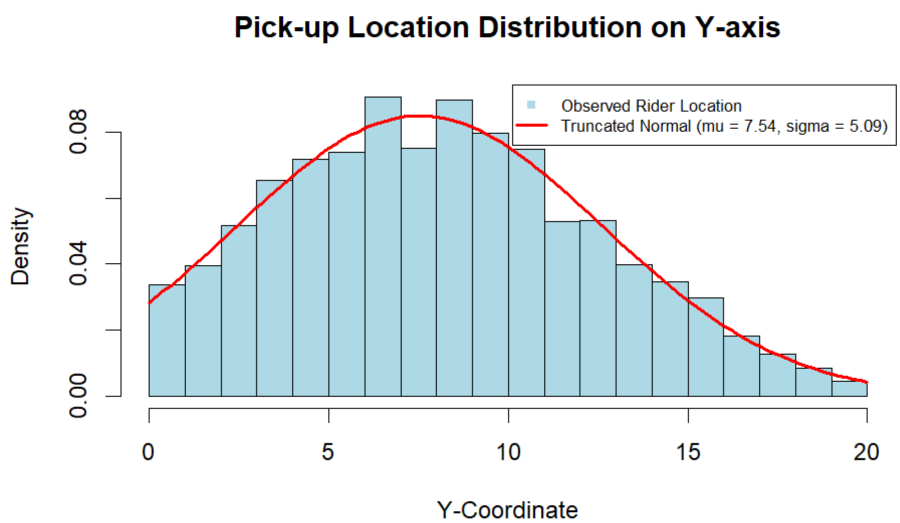

Figure 10: Observed data and Truncated Normal distribution for rider pick-up on the Y-coordinate

Similarly to driver arrivals, these results indicate that pick-up locations are not evenly distributed across the city. Instead, certain areas, corresponding to the means of the fitted normal distributions, will have a higher concentration of ride requests.

3.2.6 Rider Drop-Off location

Similar to the pick-up locations, the drop-off locations of riders are not uniformly distributed across the city. A Chi-Squared test has been performed to assess uniformity, resulting in the rejection of the null hypothesis (H_0), confirming that also drop-off locations are concentrated in specific areas rather than being equally likely across Squareshire.

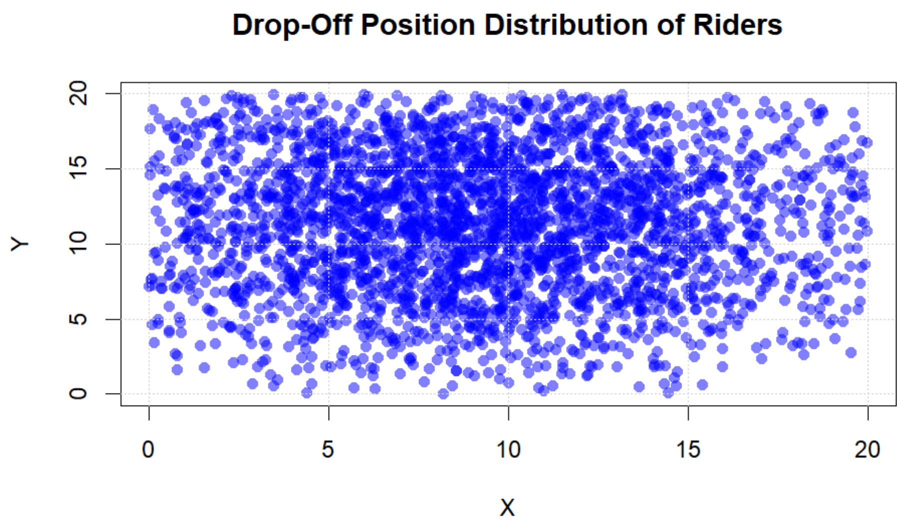

Figure 11: Scatter plot of observed drop-off locations.

The distribution has been further analysed by separately testing the X and Y coordinates. A truncated normal distribution has been fitted to both components, with parameters estimated using MLE estimators. The KS test failed to reject the truncated normal assumption, indicating a good fit for both axes.

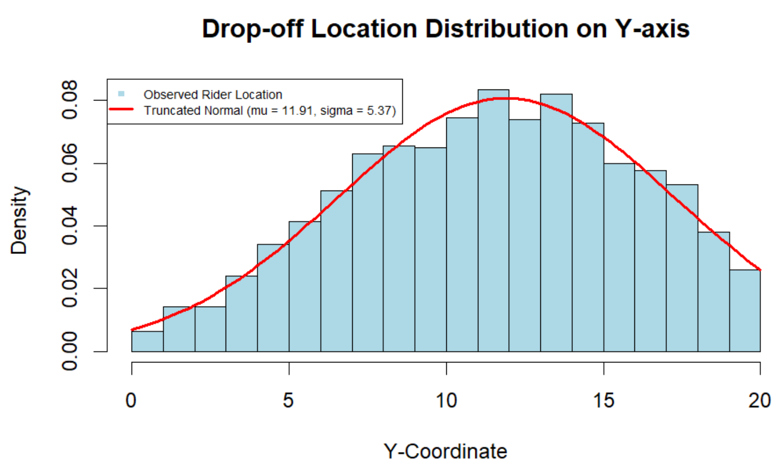

Figure 12: Observed data and Truncated Normal distribution for rider drop-off on the X-coordinate

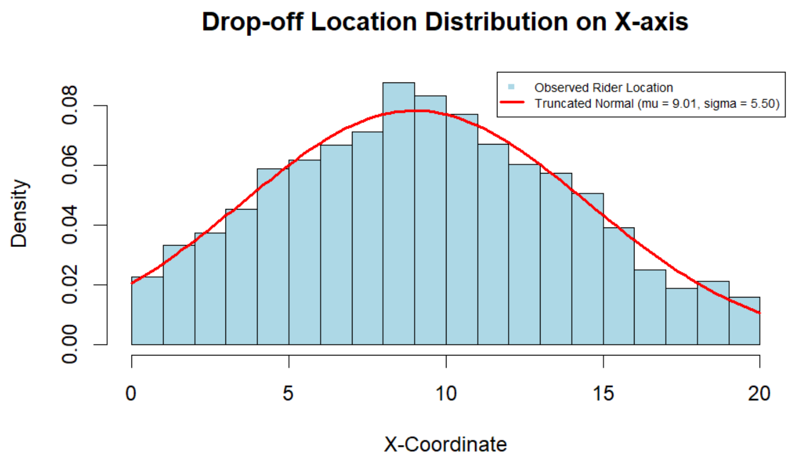

Figure 13: Observed data and Truncated Normal distribution for rider drop-off on the Y-coordinate

These results suggest that certain areas of Squareshire experience a higher concentration of drop-offs.

3.2.7 Ride time

The real-time duration of each ride has been calculated as the difference between the drop-off and pick-up times, provided by the data. The ride times are expected to vary around the expected ride time μ_t, introduced in Subsection 2.1.4, as a Uniform between $0.8\mu_t$ and $1.2\mu_t$.
To validate this assumption, the real ride times have been normalized by dividing them by their corresponding expected values, that have been computed using the pick-up and drop-off locations. If the assumption is correct, the normalized values should follow a uniform distribution between 0.8 and 1.2.

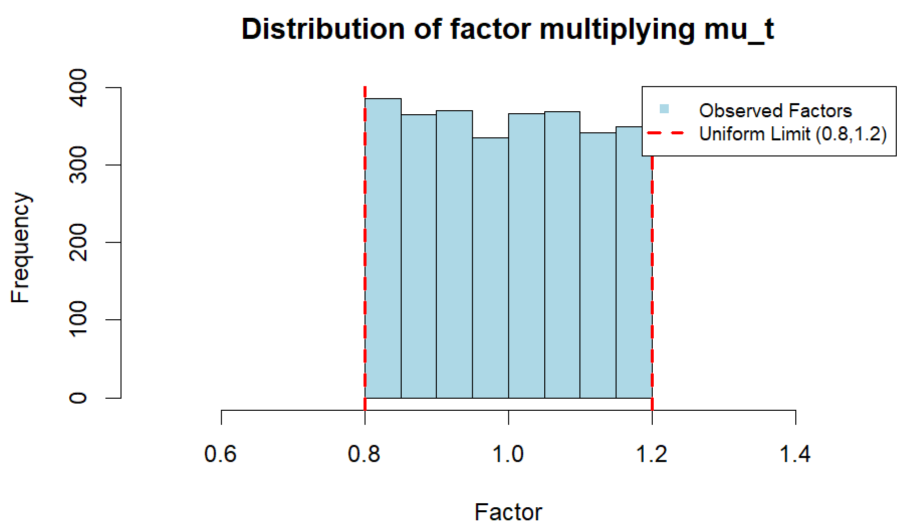

Figure 14: Factor that multiply μ_t to provide real ride times

The histogram of the normalized ride times visually supports this assumption, as the observed data is evenly spread across the expected range. A KS test has been performed, since, unlike the Chi-Squared

test, does not require independence of the data. This is particularly important in the case of ride times because they may exhibit dependencies due to factors such as traffic or time of day.
The results of the test show a p-value of 0.4059. Given that the p-value is well above the threshold of 0.1, the test fails to reject the null hypothesis, confirming that the observed ride times are consistent with the assumed uniform distribution.
This suggests that the model's approximation of trip duration variations is reasonable.

3.3 Testing conclusions

The testing process leads to significant adjustments in the simulation. The revision of driver online time to a uniform distribution between 6 and 8 hours means that drivers remain available longer than initially assumed, increasing overall driver presence in the system. The updated inter-arrival rate for the drivers of 4.09/hour results assures more driver in the system, that should reduce passenger wait times and abandonment rates. The shift from a uniform to a (truncated) normal distribution for the different locations, means that the simulation will now work under the more realistic assumption that some area in Squareshire are more busy than others. Overall, these changes will provide a more realistic basis for performance evaluation.

4 Results

An analysis has been conducted on the results of three simulations: the initial model with original distributions, a revised version using updated distributions, and a final simulation aimed at improving conditions for drivers based on previous results.
To avoid boundary issues, such as including drivers who worked for only a short time, the results (KPIs) have been computed only for drivers and riders who entered the system at least 12 hours before termination. This ensures that all included participants had enough time to complete their trips and contribute meaningful data to the analysis.
The key performance indicators presented in subsection 2.2.5 have been computed and presented in Tables 1, 2 and 3. Variability of the results, used to measure fairness among drivers, is displayed in Figure 15 and 16.

Table 1: Simulation Results for Riders

	Rate of abandonments	Average waiting time for pick up (hour)
Old dist.	0.253	0.409
New dist.	0.045	0.316
New method	0.084	0.405

Table 2: Income per hour for Drivers

	Mean	Standard deviations	0.05 quantile	0.95 quantile
Old dist.	21.062	3.353	15.834	26.411
New dist.	16.455	4.528	7.712	22.136
New method	17.773	2.567	13.353	21.939

Table 3: Resting time for Drivers

	Mean	Standard deviations	0.05 quantile	0.95 quantile
Old dist.	0.474	0.652	0.004	1.861
New dist.	2.159	1.394	0.455	4.918
New method	1.054	0.718	0.118	2.435

4.1 Old distributions and new distributions methods

The comparison between using the original and new distributions shows how the performance of the system changes, under more realistic assumptions.
Passenger satisfaction increased, with fewer ride cancellations, that go down from 25.3% to 4.5%, and shorter waiting times, from 24.5 to 19 minutes (see Table 1). This happens because drivers appear more often and stay online longer, arrival rate is 4.09 per hour instead of 3, and online time is between 6 and 8 hours instead of 5 and 8.
However, with more drivers available, average earnings dropped from £21.06 to £16.45 per hour, and incomes became more unfair: 5% of drivers now earn less than £7.71, compared to £15.83 before (Table 2). At the same time, average rest periods increased from 0.47 to 2.16 hours (Table 3), because more drivers have to share the same number of riders as before.
Overall, the service for passengers has improved significantly. However, while income fairness among drivers has increased, the average earnings have dropped considerably. In particular, the lowest 5% of drivers now earn a very low wage, which raises concerns about financial satisfaction of the drivers. Further improvements are needed to ensure a fairer distribution of income, while maintaining the benefits of the new system.

4.2 New method

The implementation of the new passenger and driver matching function (match.P.to.D), presented in subsection 2.3, considers both distance and driver working time, when performing the match. This is done through a score function. This new method has led to significant changes in system performance compared to the previous model (the new distributions method).
The new proposition prioritizes drivers who have been working less than others, ensuring a more balanced workload distribution, that was our main problem of the last simulation, since the earnings of some drivers were too low. The decision, after testing, has been of working with a value of weight equal to 10.
As a result, driver income has become more evenly distributed, with the lowest earning drivers seeing an increase in their income per hour (5% quantile goes up to £13.5 from £7.712). The average hourly income has risen from £16.45 to £17.77, while the standard deviation has decreased, increasing fairness. Additionally, drivers now experience less idle time, as resting hours have decreased from 2.16 to 1.05 hours, that provides a better utilization of driver that work on shift between 6 and 8 hours. The new resting time is more reasonable than the previous one, counting that breaks are made during different moments of the shift.
However, these improvements for drivers come at a slight cost to rider experience, as average wait times have increased from 19 to 24 minutes, but more importantly the abandonment rate has risen from 4.5% to 8.4%. This suggests that prioritizing workload over proximity in the matching process may lead to delays in assigning some passengers. While the system is now more fair for drivers, some problems arise about rider's satisfaction.

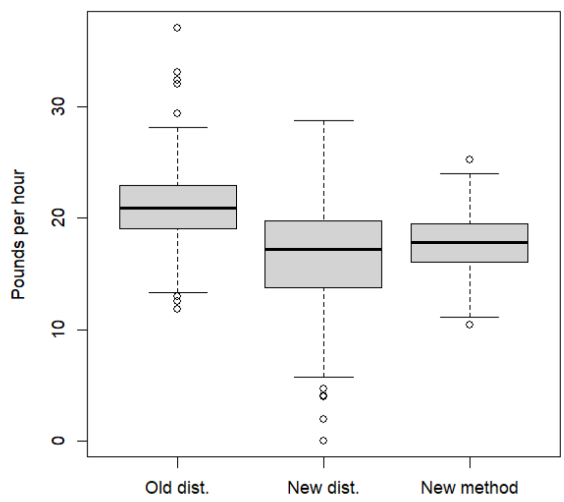

Figure 15: Boxplot for income per hour for drivers

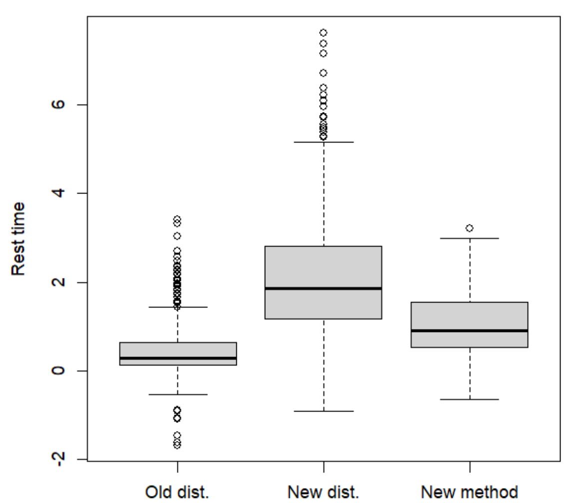

Figure 16: Boxplot for resting time for drivers

5 Conclusion

The report has analysed the performance of BoxCar's system using a series of different simulations. First, the system has been modelled using the original distributions provided by the company. Then, real-world data has been used to test these distributions, leading to a more accurate representation of reality.

The model with new distributions shows unfairness among drivers' earnings and resting time. To solve

these issues, a new driver-passenger matching method has been introduced, prioritizing both distance and driver who have worked less. This change has helped to distribute earnings more fairly and reduced excessive rest time, but slightly increased wait times and abandonment rates for riders.
The results highlight the trade-offs between optimizing rider experience and ensuring fairness for drivers.